NEW YORK UNIVERSITY
INSTITUTE OF MATHEMATICAL SCIENCES
LIBRARY
25 Waverly Place, New York 3, N.Y.

NEW YORK UNIVERSITY
INSTITUTE OF MATHEMATI [illegible] CES
LIBRARY
4 Washington Place, New York 3, N. Y.

NEW YORK UNIVERSITY
WASHINGTON SQUARE COLLEGE OF ARTS AND SCIENCE
MATHEMATICS RESEARCH GROUP

RESEARCH REPORT No. BR-3

ALGEBRAIC ASPECTS IN THE THEORY OF SYSTEMS OF LINEAR DIFFERENTIAL EQUATIONS

by

WILHELM MAGNUS

CONTRACT No. AF-18(600)-367

New York University

Washington Square College of Arts and Science

Mathematics Research Group

Research Report No. BR-3

ALGEBRAIC ASPECTS IN THE THEORY OF SYSTEMS OF LINEAR DIFFERENTIAL EQUATIONS

by

WILHELM MAGNUS

NEW YORK UNIVERSITY
INSTITUTE OF
25
SCIENCES

Wilhelm Magnus.
Wilhelm Magnus

Morris Kline
Morris Kline
Project Director

The research reported in this document has been made possible through support and sponsorship extended by the Office of Scientific Research, H.Q., Air Research and Development Command, U.S.A.F., Baltimore, Maryland, under Contract No. AF-18(600)-367. It is published for technical information only, and does not necessarily represent recommendations or conclusions of the sponsoring agency.

June, 1953

Abstract

The connection between the Baker-Hausdorff formula and systems of linear differential equations is discussed. On the algebraic side, a proof for a theorem by K. O. Friedrichs is given which characterizes Lie-elements in a free non-commutative ring. On the analytic side, criteria for the existence of solutions of systems of linear differential equations in terms of a finite number of quadratures are derived. The range of validity for the representation of the solutions in the form of an exponential function of a matrix is discussed. The case corresponding to a second-order differential equation is studied in detail.

Table of Contents

Page

Introduction 2

First Part: Formal Algebra

- I. Preliminaries 5
- II. A Theorem of Friedrichs 8
- III. Differentiation and Differential Operators 12
- IV. Zassenhaus' Formula 19

Second Part: Matrices

- V. Integration of Systems of Ordinary Differential Equations by Elementary Formulas 22
- VI. Conditions for the Existence of a Solution $Y = \exp\Omega$ for $\dot{Y} = AY$. 28
- VII. The Example of the Ordinary Differential Equation of Second Order 32

References 35

Introduction. The natural generalization of the first-order linear differential equation

$$\frac{dy}{dt} = a(t)y \tag{1}$$

is a system of n linear differential equations

$$\frac{dy_\nu}{dt} = \sum_{\mu=1}^{n} a_{\nu,\mu}(t)y_\mu; \qquad (\nu = 1,2,\ldots,n) \tag{2}$$

which is characterized by a matrix $A(t)$ with elements $a_{\nu,\mu}$ which are functions of t. This system is equivalent to a differential equation for a matrix Y:

$$\frac{dY}{dt} = A(t)Y. \tag{3}$$

The columns of Y form a set of n linear independent solutions of (3), i.e., a set of vectors with components y_ν satisfying (2). Equation (3) may be interpreted in an even more general way if we assume that Y and A are linear operators in a Hilbert space which depend on a parameter t in a manner which allows differentiation and integration with respect to t.

The importance of equations of type (3) for various branches of applied mathematics and theoretical physics is well known. The present investigation has been stimulated by two recent papers. In the first, K. O. Friedrichs [1] arrived at some purely algebraic problems in connection with the theory of linear operators in quantum mechanics. The answer to one of these problems turns out to be closely connected with the so-called Baker-Hausdorff formula which is for (3) the proper generalization of the well-known solution of (1). In another connection, H. B. Keller and J. B. Keller [2] resumed and continued the work by H. F. Baker [3] on systems of ordinary linear differential equations. In the present report their results are supplemented by a discussion of certain cases where an integration of a system by a finite number of quadratures is possible but where the matrix A in (3) cannot be diagonalized.

Algebraic questions arise in connection with (3) for the following reason: If it should happen that for any values of s,t, the "commutator"

$$\triangle_1 = A(s)A(t)-A(t)A(s) = 0, \tag{4}$$

then indeed (3) would have the solution

$$Y = \exp \int_0^t A(\tau)d\tau , \tag{5}$$

where the exponential function of a matrix is explained by the power series and where the integral of a matrix is obtained by integrating every element of it. Since (4) will not be satisfied except for particular cases, one may expect to be able to write the solution as

$$Y = \exp \Omega (t) \tag{6}$$

where Ω can be expressed in terms of $\int A d\tau$ and integrals which involve Δ_1 and commutators derived from Δ_1 and A; the commutators are derived by repetitions of the process which leads from A to Δ_1. A formula of this kind indeed exists and it is based upon the addition theorem of the exponential function for non-commutative variables. This theorem, discovered by Baker[4] and Hausdorff[5], is a formula referring to an abstract ring and hence it does not involve questions of convergence. Therefore, a formula of type (6) can be established in a formal way; however, its range of validity must then be investigated separately.

Feynman[6], using a symbolic interpretation for Eq. (5), has derived the solution of (3) in the same infinite series form obtained when (3) is solved by iteration.

The first part of this report deals with algebraic problems. Section I contains the necessary definitions and restatements of basic results. Section II contains a proof of Friedrich's theorem which characterizes Lie-elements of a free non-commutative ring in a simple manner. In Section III the continuous analogue of the Baker-Hausdorff formula is proved and a slightly simplified derivation of the formula itself is given. Section IV contains a method of computing the terms of Zassenhaus' formula which is in some respects the inverse of the Baker-Hausdorff formula.

In the second part of this report the results of the first part are applied to systems (3) of linear differential equations. Section V contains a discussion of the case where the infinite series, which constitutes the continuous analogue of the Baker-Hausdorff formula, terminates after a finite number of terms. In this case, (3) can be integrated by a finite number of quadratures; this is not explicitly clear from a consideration of the matrizant (i.e. the solution obtained by iteration). It is also shown that the conditions for the termination of the series are weaker than one might expect them to be. For instance, (5) is a solution of (3) if

$$A(t) \int_0^t A(s)ds - \left(\int_0^t A(s)ds \right) A(t) \equiv 0, \tag{7}$$

and (7) does not imply (4), at least not in the case of a matrix A which is not analytic in t. In Section VI the possibility of solving (3) in terms of an Ω defined by (6) is discussed. The following result will be proved: Let Ω_0 be a matrix in the space of the n^2 complex variables which constitute the general matrix Ω. Then a full neighborhood of the matrix Ω_0 will be mapped upon a full neighborhood of $Y_0 = \exp \Omega_0$ if and only if the difference between any two eigenvalues is not $2n\pi i$, $n = 1,2,3, \ldots , n \neq 0$. From this it follows that, in general, a differentiable Ω such that (6) is a solution of (3) will exist only in the small. The situation is investigated more closely in Section VII, where the equation (3) is interpreted as corresponding to a second-order differential equation $y' + Qy = 0$, $Q > 0$.

First Part: Formal Algebra

I. Preliminaries

A free associative ring R_o with n free generators $x_1, \ldots, x_n$ and an identity 1 will be defined by the following four axioms:

a. The elements of a field f_o of characteristic zero (for instance the field of real numbers) are in R_o, and the unit element 1 of f_o is the identity of R_o. The field f_o belongs to the center of R_o, i.e., all its elements commute with all the elements of R. We shall call f_o the field of the coefficients.

b. The addition in R_o is commutative and associative, and the multiplication is associative.

c. There exist no relations between the elements of R except those which follow from a and b.

d. Every element of R_o can be obtained from the elements of f_o and from $x_1, \ldots, x_n$ by carrying out a finite number of additions and multiplications.

It follows from the axioms a - d that the identity 1 and the products of any number of "factors" x_ν^α, i.e., the products

$$(1) \qquad x_{\nu_1}^{\alpha_1} x_{\nu_2}^{\alpha_2} \ldots x_{\nu_m}^{\alpha_m}, \qquad (\alpha_1, \alpha_2, \ldots, \alpha_m = 1,2,3, \ldots),$$

with

$$(2) \qquad \nu_1 \neq \nu_2 \neq \nu_3 \ldots \nu_{m-1} \neq \nu_m, \; \nu_\mu = 1,2, \ldots n \text{ for } \mu = 1, \ldots, m,$$

form a basis of linearly independent elements of R_o with respect to f_o.

We may extend R_o to a ring R, which consists of all the formal power series with coefficients $c_o, c(\nu_1, \ldots, \nu_m, \alpha_1, \ldots, \alpha_m)$ in f_o, the generic element A of R being

$$(3) \quad A = c_o + \sum c(\nu_1, \ldots, \nu_m; \alpha_1, \ldots, \alpha_m)\, x_{\nu_1}^{\alpha_1} \ldots x_{\nu_m}^{\alpha_m}.$$

The sum in (3) is taken over all possible combinations of integers $\nu_1,\ldots,\nu_m$, $\alpha_1,\ldots\alpha_m$ satisfying (1) and (2) for $m = 1,2,3,\ldots$. A power series of the type described by (3) will also be called a function of $x_1,\ldots,x_n$ and will be denoted by $F(x_1,\ldots,x_n)$. Problems of convergence do not play a role; if, for instance, f_o is the field of rational numbers, both

$$\sum_{n=0}^{\infty} n!\; x_1^n$$

and its square

$$\sum_{n=0}^{\infty} \sum_{r=0}^{n} r!(n-r)!\;\; x_1^n$$

belong to R. Here we have used the natural notation according to which $x_\nu^o = 1$ for any ν.

As an example which will be used later we may consider the case of two free generators x,y; we take for f_o the field of real numbers. The exponential function is defined by

$$e^x = \sum_{n=0}^{\infty} x^n/n! \tag{4}$$

and we have

$$e^x\, e^y = \sum_{n,m=o}^{\infty} x^n y^m/(n!m!). \tag{5}$$

We can find a function z of x and y such that

$$e^x\, e^y = e^z \tag{6}$$

and

$$z = u - \frac{1}{2}u^2 + \frac{1}{3}u^3 \mp \ldots , \tag{7}$$

where

$$u = (e^x e^y - 1) = x + y + \frac{x^2}{2!} + \frac{xy}{1!1!} + \frac{y^2}{2!} + \ldots . \tag{8}$$

If we substitute for u its value from (8), it is easily shown that the series in (7) leads to an element of R. For this purpose, we shall call

$$\alpha_1 + \alpha_2 + .. + \alpha_n$$

the degree (in $x_1,\ldots x_n$) of the basic element (1); in (5) the terms of degree ℓ (in x,y) will then consist of the sum

$$(9) \qquad \frac{1}{\ell!}\sum_{r=0}^{\ell}\binom{\ell}{r} x^r y^{\ell-r} \quad .$$

Only a finite number of powers of u will involve terms of a given degree in x,y, since the terms of lowest degree in u^k are of degree k in x and y.

Therefore, z becomes a power series in x,y with rational coefficients, the first terms being

$$(10) \qquad z = x+y+\frac{1}{2}\,xy-\frac{1}{2}\,yx \;+\ldots\;.$$

Baker[4] and Hausdorff [5] proved a theorem about the terms of higher degree in the power series (10) for z; we may state this theorem as follows:

Let u,v be any elements of R. Then the bracket-product or Lie-product $[u,v]$ of u and v is defined by

$$(11) \qquad [u,v] = uv-vu \;.$$

Using Lie-multiplication, we can define recursively a Lie-element of R. We shall call x and y (or, in the general case, the free generators x_ν) Lie-elements of degree one. Any linear combination of Lie-elements of degree one with coefficients from f_o and any Lie-product of Lie-elements shall also be called a Lie-element. The total set of Lie-elements obtained in this manner will be called Λ. For properties of this set see [7], [8], and [9]. In the case of two free generators x,y the general Lie-element involving terms of a degree $\leqq 3$ is

$$(12) \; c_1x + c_2y + c_{12}(xy-yx) + c_{121}\{(xy-yx)x - x(xy-yx)\}$$
$$+\, c_{122}\{(xy-yx)y - y(xy-yx)\},$$

where $c_1,\ldots,c_{122}$ are elements of the field f_o (e.g. rational numbers). The result of Baker [4] and Hausdorff [5] is that z in (10) is a Lie-element; also, a method of computing z up to terms of any fixed degree is given in their papers. Formula (6), with z given by equation (10), is called the Baker-Hausdorff formula. The connections between this formula and the theory of differential operators will be investigated in the following sections.

II. A Theorem of Friedrichs*

In a discussion of the theory of operators of quantum mechanics, Friedrichs[1] found a characterization of Lie-elements and proved it in a particular case by using a theory of representation of these operators. We shall formulate and prove Friedrichs' theorem in the case where the ring R has two free generators x,y. But the proof for a free ring with any number of generators is almost literally the same.

We construct first an isomorphic replica R' of R which has two free generators x',y'. Then we construct the direct product $\tilde{R}$ of R and R', identifying the elements of the underlying isomorphic fields of coefficients in R and R'. The ring $\tilde{R}$ is generated by x,y,x',y', where the generators are not entirely free but satisfy the relation

(1) $$xx' = x'x\ ,\ yy' = y'y,\ \ xy' = y'x,\ x'y = yx' .$$

We may consider any element of R as an element of $\tilde{R}$, which contains R. Now we can state Friedrichs' theorem:

Theorem I:

Let F(x,y) be a function of x and y. Then F is a Lie-element of R if and only if

(2) $$F(x+x',\ y+y') = F(x,y) + F(x',y') .$$

To prove this theorem, we may confine ourselves to the case where F is homogeneous in x,y, that is, where F is a sum of terms of a fixed degree It is easily seen that (2) holds if F is a Lie-element. In this case, F must be a sum of terms which have been derived from the generators by a repeated application of Lie-multiplication. Consider a term w involved in F, which has been obtained from two Lie-elements u and v of R by Lie-multiplication:

$$w = [u,v] .$$

Then u and v are of lesser degree than w and we may assume that, as functions of x and y, they satisfy (2). Since it follows from (1) that the bracket product of a factor depending on x,y and a factor depending on x',y' is always zero, we find that w also satisfies (2).

To show that only Lie-elements satisfy (2) we introduce first the following definition:

---

* A proof different from the one given here and based on a lemma due to Birkhoff and Witt [8] has been communicated to the author by Dr. P. M. Cohn of Manchester University.

Any r elements $(r=1,2,3,\ldots)$ of R are called algebraically independent if they generate a subring of R which is isomorphic with a free ring of r free generators. More specifically, let $u_1\ldots,u_r$ be the r elements of R and let $y_1,\ldots,y_r$ be free generators of a free ring R^*. Then we shall call the $u_\rho,(\rho=1,\ldots,r)$ algebraically independent if the mapping

$$y_\rho \rightarrow u_\rho \qquad (\rho = 1,\ldots,r)$$

determines an isomorphic (one-to-one) correspondence between R^* and the smallest subring of R which contains the u_ρ. We define the maps of sums and products in the natural manner.)

Next, we need the following two lemmas, which have been proved elsewhere (see [7]):

Lemma 1. Let $u_1,\ldots,u_r$ and v be $r+1$ algebraically independent elements of a free ring R^*. For $\ell = 1,2,3,\ldots$ and $\rho = 1,2,\ldots r$, let

$$u_\rho^{(\ell)} = \Big[\big[\ldots\big[[u_\rho v]\, v\big]\ldots\big]v\Big], \tag{3}$$

$$u_\rho^{(o)} = u_\rho \,. \tag{4}$$

The right-hand side in (3) is the result of an ℓ-fold bracket-multiplication of u_ρ by v from the right. Then all the $u_\rho^{(\ell)}$ are algebraically independent (for $\ell=0,1,2,\ldots$) .

Lemma 2. Every function $H(u_1,\ldots,u_r,v)$ can be written in a unique way in the form

$$H = H_o + H_1 v + H_2 v^2 + \ldots, \tag{5}$$

where $H_o,H_1,H_2,\ldots$ are functions of the $u_\rho^{(\ell)}$, $(\ell=0,1,2,\ldots)$.

We shall apply these lemmas to the proof of Friedrichs' theorem. Let $F(x,y)$ be a function which satisfies (2) and is of degree d in x,y. F is then expressed in terms of Lie-elements of the first degree.

Assume that F could also be written as a function $H(u_1,\ldots,u_r,v)$ of algebraically independent Lie-elements of degrees

$$\ell_1 \geqq \ell_2 \geqq \ldots \geqq \ell_r \geqq \ell_o \,,$$

where ℓ_o is the degree of v in x,y. We shall derive the fact that then F can also be written as a function of algebraically independent Lie-elements $u_r^{(\ell)}$ of degree $\ell_r + \ell\ell_o$, and once this has been established it is clear that F must be a Lie-element. From Lemma 2 we see first that the degree d of F is not less than the value of the largest of the numbers $\ell_1, \ldots, \ell_r$, ℓ_0 and second, that the transition from the set of variables u_r and v to the set $u_r^{(\ell)}$ eliminates one of the variables of lowest degree ℓ_0. After a finite number of transitions of this type, we will arrive at a set of variables which are Lie-elements and none of which is of a degree less than d. Then F must be a linear combination of these variables, which proves that it must be a Lie-element.

Therefore, only the following remains to be shown: If, in terms of the Lie-elements $u_1, \ldots, u_r$, v,

$$F(x,y) = H(u_1, \ldots, u_r, v), \tag{6}$$

and if F has the property (2), then

$$F(x,y) = H_0(u_\rho^{(\ell)}) . \tag{7}$$

Since the "variables" $u_1, \ldots, u_r$, v are Lie-elements of the ring R generated by x,y, we have

$$w(x+x', y+y') = w(x,y) + w(x',y') , \tag{8}$$

where w stands for any one of the elements $u_1, \ldots, u_r$, v. We shall write w' for $w(x',y')$ and, correspondingly, u_ρ', v' for $u_\rho(x',y')$, $v(x',y')$. Now we have from (2), (6) and (8)

$$F(x+x', y+y') = H(u_1+u_1', \ldots, u_r+u_r', v+v') . \tag{9}$$

Applying Lemma 2 to H, we find:

$$\begin{aligned} & H(u_1+u_1', \ldots, u_r+u_r', v+v') \\ & \quad = H_0(u_\rho^{(\ell)} + u_\rho^{(\ell)'}) + H_1(u_\rho^{(\ell)} + u_\rho^{(\ell)'})(v+v') \\ & \qquad + H_2(u_\rho^{(\ell)} + u^{(\ell)'})(v+v')^2 + \ldots , \end{aligned} \tag{10}$$

where the $u_\rho^{(\ell)'}$ are derived from the $u_\rho^{(\ell)}$ by the transition from x,y to x',y'.

Now condition (2) gives

$$(11)\quad H(u_1+u_1',\dots,u_r+u_r',v+v')$$

$$= H(u_1,\dots,u_r,v) + H(u_1',\dots,u_r',v')$$

$$= H_0(u_\rho^{(\ell)}) + H_0(u_\rho^{(\ell)\prime})$$

$$+ H_1(u_\rho^{(\ell)})\,v + H_1(u_\rho^{(\ell)\prime})\,v'$$

$$+ H_2(u_\rho^{(\ell)})\,v^2 + H_2(u_\rho^{(\ell)\prime})\,v'^2 + \dots .$$

According to Lemma 2, the coefficients of the products $v^k\,v'^{k'}$ on the right-hand sides of (10) and (11) must be the same. From this we have

$$(12)\qquad H_0(u_\rho^{(\ell)}+u_\rho^{(\ell)\prime}) = H_0(u_\rho^{(\ell)}) + H_0(u_\rho^{(\ell)\prime})$$

$$(13)\qquad H_1(u_\rho^{(\ell)}+u_\rho^{(\ell)\prime}) = H_1(u_\rho^{(\ell)}) = H_1(u_\rho^{(\ell)\prime})$$

$$(14)\qquad H_2(u_\rho^{(\ell)}+u_\rho^{(\ell)\prime}) = H_2(u_\rho^{(\ell)}) = H_2(u_\rho^{(\ell)\prime})$$

. . . .

From equations (13) and (14) we may conclude that $H_1, H_2, \dots$ vanish identically in the $u_\rho^{(\ell)}$, and equation (12) verifies that H_0 has the same property as the properties of F from which we started. This finishes the proof of Theorem I. It has been used by Friedrichs to prove the following

Theorem II . Let x,y be free generators of an associative ring R. Let z and w be the elements defined by

$$(15)\qquad e^x e^y = e^z$$

$$(16)\qquad e^{-x}\,y\,e^x = w .$$

Then z and w are Lie-elements in R.

The proof follows immediately from Theorem I if we observe that $e^{x+x'} = e^x \cdot e^{x'}$ and $xx' = x'x$. Baker[4] and Hausdorff[5] proved Theorem II by a recursive construction of z and by an explicit formula for w. Once one knows that w is a Lie-element, it is easy to determine it explicitly. Since it is of first degree in y, it must be expressible in terms of the Lie-elements

$$\{y,x^{\ell}\} = [\,..[\,[y,x]x\,]...x\,] \tag{17}$$

which are obtained from y by an ℓ-fold bracket multiplication by x. The result is

$$w = \sum_{\ell=0}^{\infty} \frac{1}{\ell!}\{y,x^{\ell}\}, \tag{18}$$

where $\{y,x^{\ell}\}$ stands for y itself.

III. Differentiation and Differential Operators

Let R be a free associative ring with free generators $x,y,x_1,y_1 \ldots$ Let F(x,y) be an element of R and let λ be a parameter, i.e., an arbitrary number from the field f_0 (for instance an arbitrary real number). Then we can expand $F(x+\lambda x_1,y)$ in a series of powers of λ :

$$F(x+\lambda x_1\,,\,y) = F(x,y) + \lambda F_1(x,x_1,y) + \ldots\,. \tag{1}$$

We shall call the coefficient of λ a derivative and write

$$F_1(x,x_1,y) = (x_1 \frac{\partial}{\partial x})F(x,y)\;. \tag{2}$$

The word "derivative" was introduced by Hausdorff. Another more customary notation for this coefficient is "polar", and the process by which F_1 is obtained from F is also called "polarization". The definition of a derivative used here is different from the one used by Falk [10]. If F is a monomial (that is, an element of the type (I.1)), polarization consists of first replacing one factor x by x_1 in every possible way and then adding all the resulting terms afterwards. The polar of a sum of terms is the sum of the polars of the terms. If F does not involve x, its polar with respect to x is zero. It is not necessary that x_1 be different from x; we can define $(x\frac{\partial}{\partial x})F$ by forming $(x_1\frac{\partial}{\partial x})F$ and substituting x for x_1 after the polarization. We may also substitute any element u of R for x_1 after polarization and will denote the result by $(u\,\frac{\partial}{\partial x})\,F$.

We shall need the following formulas which are due to Hausdorff [5]:

$$\begin{aligned} \left((y\tfrac{\partial}{\partial x})e^{x}\right)e^{-x} &= y - \tfrac{1}{2!}[y,x] + \tfrac{1}{3!}\left[[y,x]x\right] - \tfrac{1}{4!}\left[\left[[y,x]x\right]x\right] + \dots \\ e^{-x}(y\tfrac{\partial}{\partial x})e^{x} &= y + \tfrac{1}{2!}[y,x] + \tfrac{1}{3!}\left[[y,x]\,x\right] + \tfrac{1}{4!}\left[\left[[y,x]\,x\right]x\right] + \dots\,. \end{aligned} \tag{3}$$

Since we shall need repeated bracket multiplication in various formulas, we shall simplify the notation by writing

$$\left\{y,\ P(x)\right\} = p_o y + p_1[y,x] + p_2\left[[y,x]\ x\right] + \dots \tag{4}$$

if

$$P(x) = p_o + p_1 x + p_2 x^2 \tag{5}$$

is a power series in x. We may then write the equations (3) in the form

$$\begin{aligned} e^{-x}\,(y\tfrac{\partial}{\partial x})e^{x} &= \left\{y,\ \frac{e^{x}-1}{x}\right\} \\ \left((y\tfrac{\partial}{\partial x})e^{x}\right)e^{-x} &= \left\{y,\ \frac{1-e^{-x}}{x}\right\}\,. \end{aligned} \tag{6}$$

If $P(x)$ has a constant term $p_o \neq 0$, there exists a power series $Q(x)$ such that

$$P(x)\ Q(x) = 1\,. \tag{7}$$

We have then, according to Hausdorff [5] :

Lemma 3 : Let $P(x)$, $Q(x)$ satisfy (7) and let

$$\left\{y,\ P(x)\right\} = u. \tag{8}$$

Then

$$y = \left\{u, Q(x)\right\}. \tag{9}$$

The proof consists of a simple straightforward computation, using the relations between the coefficients of $P(x)$ and $Q(x)$ which result from (7).

We shall consider now differentiable functions of a real parameter t in an associative ring R. Examples of functions of this type are given by finite or infinite matrices with real or complex elements which are differentiable functions of t. For our purposes, two types of properties of these differentiable functions are required. As formal properties we need these:

(a). A(t) is an element of R for all real values of t. For any sufficiently small ε,

$$A(t+\varepsilon) = A(t) + \varepsilon A_1(t) + \varepsilon^2 A_2(t) + \dots , \tag{10}$$

where the terms on the right-hand side of (10) are in R and where their sum shall be defined in R. We shall define $\frac{dA}{dt}$ to be $A_1(t)$.

(b). The formal laws of differentiation for a sum and a product hold.

(c). If P(x) is a power series in x and if $P(A(t)) = B(t)$ exists in R, then

$$\frac{dB}{dt} = \left(y \frac{\partial}{\partial x}\right) P(x) , \qquad y = \frac{dA}{dt} , \quad x = A(t) . \tag{11}$$

(d). Given A(t), there exists in R a uniquely determined function $A^*(t)$ such that

$$\frac{dA^*}{dt} = A(t) , \quad A^*(0) = 0 . \tag{12}$$

We shall write

$$\int_0^t A(\tau)d\tau = A^*(t) . \tag{13}$$

A second type of property, which is more difficult to describe, concerns problems of convergence. For the present purposes it will be necessary to assume that exp A exists for all functions under consideration and is differentiable in the sense that (13) exists. Also, it will be necessary to assume that certain repeated integrals (where integration is defined by (12),(13)) exist.

It may suffice here to make two remarks. Firstly, the definition of a free ring R is already based implicitly on an assumption about convergence in the sense that any sequence u_n $(n = 1,2,3,\dots)$ of elements of R has the limit zero if minimum degrees μ_n of terms involved in u_n satisfy

$$\lim_{n \to \infty} \mu_n^{-1} = 0. \tag{14}$$

Secondly, we can construct an example where all the functions considered have the convergence properties which are required in this section. Let R be a free associative ring with a finite number of free generators. Let u_n be any sequence of elements of R for which the minimum degrees μ_n of terms involved in u_n have the property (14). Then for any choice of the u_n,

$$A(t) = \sum_{n=0}^{\infty} u_n t^n$$

has all the convergence (or existence) properties needed in this section.

The case where the A(t) are finite matrices with elements depending on t will be considered extensively in the next section.

We can now state

Theorem III . *Let A(t) be a known function of t in an associative ring R, and let U(t) be an unknown function satisfying*

$$\frac{dU}{dt} = AU , \qquad U(o) = 1. \tag{15}$$

Then, if certain unspecified conditions of convergence are satisfied, U(t) can be written in the form

$$U(t) = \exp \Omega (t) \tag{16}$$

where

$$\frac{d\Omega}{dt} = \left\{A, \frac{\Omega}{e^{\Omega}-1}\right\} = \sum_{n=0}^{\infty} \beta_n \left\{A, \Omega^{\ n}\right\}$$

(17)

$$= A - \frac{1}{2}[A,\Omega] + \frac{1}{12}\Big[[A,\Omega]\Omega\Big] - \frac{1}{4!} \cdot \frac{1}{30}\Big[\big[[[A,\Omega]\Omega]\Omega\big]\Omega\Big] \pm \dots .$$

The β_n vanish for $n = 3,5,7, \ldots,$ and $\beta_{2m} = (-1)^{m-1} B_{2m}/(2m)!$, where the B_{2m} (for $m=1,2,3\ldots$) are the Bernoulli numbers. Integration of (17) by iteration leads to an infinite series for Ω the first terms of which (up to terms involving three integrations) are:

$$\Omega = \int_0^t A(\tau)\, d\tau - \frac{1}{2}\int_0^t \left[A(\tau), \int_0^\tau A(\sigma)d\sigma\right] d\tau$$

$$+ \frac{1}{4}\int_0^t \left[A(\tau), \int_0^\tau \left[A(\sigma), \int_0^\sigma A(\rho)d\rho\right] d\sigma\right] d\tau$$

$$+ \frac{1}{12}\int_0^t \left[\left[A(\tau), \int_0^\tau A(\sigma)d\sigma\right] \int_0^\tau A(\sigma)d\sigma\right] d\tau$$

$$+ \ldots.$$

Formula (18) is the continuous analogue of the Baker-Hausdorff formula by which z is expressed in (II.15) as a Lie-element in terms of x, y. The first terms of the Baker-Hausdorff formula are*

$$z = x + y + \frac{1}{2}[x,y] + \frac{1}{12}[[x,y]\,y] - \frac{1}{12}[[x,y]\,x]$$

$$+ \frac{1}{24}[\,[[x,y]\,y] + \frac{1}{720}[[[[x,y]\,x]\,x]\,x]$$

$$- \frac{1}{720}[[[[x,y]\,y]\,y]\,y] + \frac{1}{180}[[[[x,y]\,x]\,x]\,y]$$

$$- \frac{1}{180}[[[[x,y]\,x]\,y]\,y] + \frac{1}{120}[[[x,y]\,x], [x,y]]$$

$$- \frac{1}{120}[[[x,y]\,y], [x,y]] + \ldots$$

---

* The Computation Laboratory of the National Bureau of Standards has provided a method of calculating the coefficients of the Baker-Hausdorff formula on its electronic computing machine. The method is based on the Dynkin-Specht-Wever formula and on the procedure indicated in the present paper. (See Section IV).

To derive (19), we can proceed in the same way by which (18) was proved. We observe first that according to (3) or (6):

(20) $$e^{-z}\left(u\frac{\partial}{\partial x}\right)e^{z} = \sum_{n=1}^{\infty}\left\{\zeta, z^{n-1}\right\}/n!\ ,$$

where

(21) $$\zeta = \left(u\frac{\partial}{\partial x}\right)z$$

and

(22) $$\left\{\zeta, z^{m}\right\} = \left[\,\left[\ldots[\zeta, z]\ldots\right]z\,\right].$$

The quantity u may be considered as a new generator or variable. Now we have from (II.15), (3), and (II.18):

(23) $$e^{-z}\left(u\frac{\partial}{\partial x}\right)e^{z} = e^{-y}\,e^{-x}\left(u\frac{\partial}{\partial x}\right)e^{x}\,e^{y}$$

(24) $$= e^{-y}\left(\sum_{n=1}^{\infty}\left\{u, x^{n-1}\right\}/n!\,\right)e^{y}\ .$$

Comparing (20) and (24) we have

(25) $$\sum_{n=1}^{\infty}\left\{\zeta, z^{n-1}\right\}/n! = e^{-y}\left\{u, \frac{e^{x}-1}{x}\right\}e^{y}\ ,$$

and (25) holds for any u if we define ζ by (21). Now we choose u in such a manner that the right-hand side in (25) becomes as simple as possible.

For this purpose we try to determine u in such a way that

(26) $$\left\{u, \frac{e^{x}-1}{x}\right\} = y.$$

This is possible and, in accordance with Lemma 3 (formulas (8) and (9)) we have

(27) $$u = \left\{y, \frac{x}{e^{x}-1}\right\}.$$

With this value of u, the right-hand side of (25) is equal to $y = e^{-y}ye^{y}$, and now, using Lemma 3 again in order to express ζ in terms of y and z, we obtain from (25)

$$\zeta = (u \frac{\partial}{\partial x}) z = \left\{ y, \frac{z}{e^{z}-1} \right\} = y - \frac{1}{2} [y,z] + \frac{1}{12} \left[[y,z] \, z \right] + \dots . \tag{28}$$

In a similar manner we find

$$\left((v \frac{\partial}{\partial y}) e^{-z} \right) e^{-z} = \left\{ \zeta^{*}, \frac{e^{z}-1}{z} \right\} = \left((v \frac{\partial}{\partial y}) \, e^{-y} \right) e^{y} = \left\{ v, \frac{e^{y}-1}{y} \right\}, \tag{29}$$

where

$$\zeta^{*} = (v \frac{\partial}{\partial y}) \, z \, . \tag{30}$$

If we choose $v = y$ we find from (29), Lemma 3, and (28), that

$$(y \frac{\partial}{\partial y}) \, z = \left\{ y, \frac{z}{e^{z}-1} \right\} = (u \frac{\partial}{\partial x}) \, z \, . \tag{31}$$

Since z is a Lie-element and since every Lie-element in x,y contains both x and y if its degree is at least two, we may write z in the form

$$z = z_{o} + z_{1} + z_{2} + \dots + z_{n} + \dots ,$$

where $z_{o} = x$ and where, for $n = 1,2,3, \dots$, the term z_{n} consists of all the terms in z which are of degree n with respect to y. Since terms of the same degree with respect to y on the two sides of an equation must be equal, we find from

$$\begin{aligned} (y \frac{\partial}{\partial y}) \, z &= z_{1} + 2z_{2} + \dots + nz_{n} + \dots \\ &= (u \frac{\partial}{\partial x}) (x + z_{1} + z_{2} + \dots + z_{n} + \dots) \end{aligned} \tag{32}$$

that

$$z_{1} = u = (u \frac{\partial}{\partial x}) \, x, \tag{33}$$

$$z_{2} = \frac{1}{2} (u \frac{\partial}{\partial x}) z_{1} , \tag{34}$$

- - - - - - - - - -

$$z_{n} = \frac{1}{n} (u \frac{\partial}{\partial x}) z_{n-1} \, . \tag{35}$$

The meaning of equations (33) to (35) is that we apply first the operator $u \frac{\partial}{\partial x}$ and then substitute for u the expression in (27). Apparently, (35) allows a computation of z by recurrence.

We can derive Theorem III from Hausdorff's formula by considering integrals as limiting cases of a sum. However, the direct proof of (18) by iterated integration of (17) seems to be simpler.

IV. Zassenhaus' Formula

Let R be the free ring with two generators x,y and with rational coefficients. It has been observed by Zassenhaus [11] that there must exist a formula which may be called the dual of Hausdorff's formula. We may state his result as follows:

There exist uniquely determined Lie-elements $C_n (n = 2,3,4,\dots)$ in R which are exactly of degree n in x,y such that

$$e^{x+y} = e^x e^y e^{C_2} e^{C_3} \dots e^{C_n} \dots \tag{1}$$

The existence of a formula of type (1) is an immediate consequence of Hausdorff's theorem. But the computation of C_n becomes rather difficult if it is based on Hausdorff's complicated formula. A simpler method for the calculation of the C_n can be derived from a result due to Dynkin [12], Specht [13] and Wever [14]. The method employed here has already been used by Dynkin to derive the coefficients of the terms of degree n in Hausdorff's formula without the use of the coefficients of terms of lower degree.

For every element F(x,y) in R we define a corresponding Lie-element $\{F\}$, where the "curly bracket operator" $\{\,\}$ has the following properties:

a. For any element C in the field of coefficients,

$$\{CF\} = C\{F\}. \tag{2}$$

b. For any two elements F_1, F_2 of R

$$\{F_1 + F_2\} = \{F_1\} + \{F_2\}. \tag{3}$$

c. Let x_ν, for $\nu = 1,2,\dots, n$, be any one of the generators. Then for any monomial $x_1 x_2 \dots x_n$ we define

$$\{x_1 x_2 \dots x_n\} = \left[\left[\dots\left[[x_1 x_2]x_3\right]\dots\right] x_n\right]; \tag{4}$$

and for the identity we define

$$\{1\} = 0 .$$

It is clear that the operator $\{\}$ is defined uniquely for all F in R by the rules a,b,c. Now the result proved in [12], [13], [14] is: Let G be a homogeneous Lie-element in R which is of degree n. Then

$$\{G\} = nG . \tag{5}$$

The following result can be derived easily from Wever's paper [14]:

Lemma 4. If G is a homogeneous Lie-element and F is any element of R, then

$$\{G^2F\} = 0. \tag{6}$$

Now we expand both sides of (1) in power series and apply the operator $\{\}$.

According to Lemma 4 we find

$$\{e^{x+y}\} = \{1\} + \{x + y\} + \frac{1}{2!}\{x + y\}^2 + \frac{1}{3!}\{x + y\}^3 + \cdots \tag{7}$$

$$= x + y$$

since $\{(x + y)^n\} = 0$ if $n > 1$. In the same manner we find for the left-hand side in (1)

$$\{e^x e^y e^{C_2} e^{C_3} \ldots\} = \{\sum x^n/n! \sum y^n/n! \sum C_2^n/n! \ldots\} \tag{8}$$

$$= x + y + \{xy\} + \{C_2\} + \{xy^2\}/2! + \{xC_2\} + \{yC_2\}$$

$$+ \{C_3\} + \ldots ,$$

where the omitted terms are of a degree greater than three. By comparing terms of the same degree in (7) and (8) we find

$$\{C_2\} + \{xy\} = 0$$

$$\{C_3\} + \{xC_2\} + \{yC_2\} + \frac{1}{2}\{xy^2\} = 0$$

$$- - - - - - - - - - - - - - - - - - ,$$

and therefore, because of (5),

$$C_2 = -\frac{1}{2}\,[x,y]$$

$$C_3 = -\frac{1}{6}\Big[\,[x,y]y\,\Big] + \frac{1}{6}\Big\{(x+y)(xy-yx)\Big\}$$

$$= -\frac{1}{3}\Big[\,[x,y]y\,\Big] - \frac{1}{6}\Big[\,[x,y]x\,\Big]\,.$$

It is clear that by this method we may also compute C_n for any $n > 3$ by recurrence formulas.

Second Part: Matrices

V. Integration of Systems of Ordinary Differential Equations by Elementary Formulas

Let $A(t)$ and $Y(t)$ be n by n matrices the elements of which depend on a parameter t. We consider the system of linear homogeneous differential equations

$$\frac{dY}{dt} = AY \tag{1}$$

subject to the initial conditions

$$Y(o) = I, \tag{2}$$

where I denotes the unit matrix.

From well-known general theorems we know that (1) always has a uniquely determined solution $Y(t)$ which is continuous and has a continuous first derivative in any interval in which $A(t)$ is continuous. The elements of the k-th column of $Y(t)$ are the solutions of y_ν of the system of linear differential equations

$$\frac{dy_\nu}{dt} = \sum_{\mu=1}^{n} a_{\nu,\mu}\, y_\mu \;, \qquad (\nu = 1,2,\ldots, n) \tag{3}$$

subject to the initial conditions

$$y_\nu(o) = 0 \text{ if } \nu \neq k, \quad y_k(o) = 1. \tag{4}$$

The $a_{\nu,\mu}$ are, of course, the elements of A, and they are functions of t. The determinant of Y is always different from zero; its value is given by

$$| Y | = \exp \left(\int_o^t \sum_{\nu=1}^{n} a_{\nu\nu} (s)\, ds \right) . \tag{5}$$

We wish to apply Theorem III and in particular formula (III.18) to equation (1), assuming that Y can be written in the form $Y = \exp \Omega$. In general, the use of Theorem III involves difficulties of convergence; some of these will be discussed in Sections VI, VII. But there is one case in which (III.18) clearly determines Ω for all values of t, namely when the series in the right-hand side of (III.18) terminates. This will happen, for instance, if

(6) $$A(t)\left(\int_0^t A(\tau)\,d\tau\right) - \left(\int_0^t A(\tau)d\tau\right)A(t) \equiv 0$$

identically for all values of t. If (6) is true, Ω becomes simply

(7) $$\int_0^t A(\tau)d\tau$$

and $Y = \exp\Omega$ satisfies (1).

In order to state a concise result, we introduce the following:

Definition of a Lie-integral Functional. Let A(t) be an integrable function of t (in the ordinary sense). We define a Lie-integral functional Φ_n of weight n of A recursively as follows:

(i) The functional of weight 1 is any multiple of

(8) $$\int_c^t A(s)\ ds.$$

(ii) Let $\bar{\Phi}_\lambda, \bar{\Phi}_\mu, \dots, \bar{\Phi}_\nu$ be any functionals of weight $\lambda, \mu, \dots, \nu$ such that

(9) $$\lambda + \mu + \dots \nu + \rho = n-1\ .$$

Then a functional of weight n is defined as any linear combination of terms of the type

(10) $$\int_c^t \left[\left[\cdots\left[\left[A(s), \bar{\Phi}_\lambda\right]\bar{\Phi}_\mu\right]\cdots\right]\bar{\Phi}_\nu\right]\bar{\Phi}_\rho\right]ds\ ,$$

where $\bar{\Phi}_\lambda, \dots, \bar{\Phi}_\rho$ are written as functions of the independent variable s.

Apparently, the terms involving 1, 2, ..., n integrations in (III.18) are functionals of the type described above; we shall call them the Baker-Hausdorff functionals B_n of A(t) (for $c = 0$), and we shall write (III.18) in the form

(11) $$\Omega = B_1 + B_2 + B_3 + \dots ,$$

where the B_n shall be written

$$B_n(A, t, c) \tag{12}$$

if the matrix A, the variable t, and the constant c are to be exhibited. In (III.18), we had assumed that $c = 0$. Now we can state

Theorem IV. If all Lie-integral functionals of A of a weight m vanish ($n < m \leqq 2n-1$), then the solution of (1) with initial conditions (2) is given by $Y = \exp \Omega$, where

$$\Omega = \sum_{\nu=1}^{n} B_\nu (A,t,0). \tag{13}$$

The B_ν are the Baker-Hausdorff functionals defined by (III.18) and (11).

A sufficient (but not a necessary) condition for the vanishing of Lie-integral functionals of weight greater than n is that

$$\Big[\Big[\cdots \big[[A(s_1), A(s_2)] A(s_3) \big] \cdots \Big] A(s_{n+1}) \Big] = 0 \tag{14}$$

for any choice of $s_1, \ldots, s_{n+1}$.

Clearly, the fact that all functionals of type (10) vanish for weight m between n and 2n-1 implies that but a finite number of linearly independent functionals must vanish identically.

In order to prove Theorem IV we need first:

Lemma 5: If all Lie-integral functionals of a weight m vanish, where $n < m \leqq 2n-1$, then all Lie-integral functionals of any weight $m > n$ also vanish.

Proof. We shall consider a functional of type (10), assuming now that $m = 1 + \lambda + \mu + \ldots + \rho$ and that $m > n$. Since our Lemma is trivial for $n = 1$, we may assume that $n > 1$. Now we shall apply induction with respect to m, assuming that $m \geqq 2n$. If one of the weights $\lambda, \mu, \ldots, \rho$ is greater than n, the corresponding Φ vanishes identically and the Lemma holds. But suppose all of the weights $\lambda, \mu, \ldots, \rho$ are $\leqq n$; then we consider the first number S greater than n in the sequence

$$1,\ 1 + \lambda,\ 1 + \lambda + \mu, \ldots,\ 1 + \lambda + \mu + \ldots + \nu,\ 1 + \lambda + \ldots + \nu + \rho.$$

This number is necessarily at most equal to 2n. If $S < 2n$, for example, if $S = 1 + \lambda + \ldots + \nu$, then

$$\left[\left[\cdots\left[[A, \Phi_\nu]\, \phi_\mu\right]\cdots\right] \Phi_\nu\right] \tag{15}$$

is the derivative of a functional of degree S, where $n < S < 2n$, and therefore not only this functional but also (10) vanishes identically. Hence for this case our Lemma is true. There remains the case $S = 2n$, which can take place only if the last term ν in $S = 1 + \lambda + \ldots + \nu$ is equal to n. Now we use a Lemma proved by Wever [14] according to which a Lie-product (15) can also be written as a sum of Lie-products in which Φ_ν always appears in the first place, but in which the factors and the arrangement of the brackets are the same as in (15). This Lemma follows without difficulty from the Dynkin-Wever-Specht formula (IV.5). Consider now any Lie-product of type (15) in which the first factor is ϕ_ν. The second factor is either A or one of the other factors, which may be called Ψ. Now we merely have to show that

$$[\Phi_\nu, A] = 0, \qquad [\Phi_\nu, \Psi] = 0 . \tag{16}$$

If (16) holds, any product involving the left-hand sides in (16) also vanishes and therefore the product in (15) vanishes. Now (16) is true because $[\Phi_\nu, A] = -[A, \Phi_\nu]$ is the derivative of a vanishing functional of weight $n + 1$. Similarly, $[\phi_\nu, \phi_\mu] = 0$ since

$$\frac{d}{dt}[\Phi_\nu, \Psi] = \left[\frac{d\Phi_\nu}{dt}, \Psi\right] - \left[\frac{d\Psi}{dt}, \Phi_\nu\right] . \tag{17}$$

Both $d\Phi_\nu/dt$ and $d\Psi/dt$ are sums of terms of the type

$$[\cdots[[A, \Psi_1]\Psi_2]\cdots] ,$$

where $\Psi_1, \Psi_2, \ldots$ are Lie-integral functionals. Therefore, the right-hand side of (17) vanishes since the individual terms are derivatives of Lie-integral functionals of a weight k ($n+1 \leqq k \leqq 2n-1$) . The inequalities for k follow from the fact that the weight of Φ_ν equals n and the weight of Ψ is less than n and at least equal to unity.

This finishes the proof of Lemma 5. Probably, a better result could be obtained for any given n; for example, for $n=2$ it is easily shown that the vanishing of all functionals of weight 3 implies the vanishing of all functionals of weight 4. For $n=1$, it is trivial that all functionals of a weight $\geqq 2$ vanish if those of weight 2 vanish.

To prove Theorem IV we observe that all steps in the formal proof of (III.18) now involve only a finite number of terms. Firstly, the solution Ω of (17) by iteration now gives a finite sum of functionals which we denote by B_ν $(\nu=1,\dots,n)$ in accordance with (11). Secondly, it follows directly that the Lie-product

$$(18) \qquad [\,[\dots[\,[\tfrac{d\Omega}{dt},\Omega]\,\Omega]\dots]\,\Omega]$$

of at least $m+1$ factors vanishes identically of $m \geqq n$. Now we derive from (III.3)

$$(19) \quad \left(\frac{d}{dt}e^{\Omega}\right)e^{-\Omega} = \frac{d\Omega}{dt} - \frac{1}{2!}\left[\frac{d\Omega}{dt},\Omega\right] + \frac{1}{3!}\left[\left[\frac{d\Omega}{dt},\Omega\right]\Omega\right] \mp \dots ;$$

This formula makes sense for any differentiable finite matrix $\Omega(t)$ since it can be shown to be an absolutely and uniformly convergent rearrangement of the series obtained by differentiating $\exp(\Omega)$ directly term by term and multiplying by $\exp(-\Omega)$ afterwards. From (18) it follows that the right-hand side in (19) is a terminating series. Calling its sum B, we find directly from (18), (19) that

$$(20) \qquad \frac{d\Omega}{dt} = \sum_{\nu=0}^{n-1} \beta_\nu \left\{B,\Omega^\nu\right\},$$

where the β_ν are explained in Theorem III. On the other hand, Ω could also have been derived from

$$(21) \qquad \frac{d\Omega}{dt} = \sum_{\nu=0}^{n-1} \beta_\nu \left\{A,\Omega^\nu\right\}$$

since the higher terms in (17) do not contribute to Ω. Putting $B-A = C$ we merely have to show that the equation

$$(22) \qquad \sum_{\nu=0}^{n-1} \beta_\nu \left\{C,\Omega^\nu\right\} = 0$$

cannot have a solution C which does not vanish identically and which is such that $\left\{C, \Omega^{m}\right\} = 0$ if $m \geqq n$. This can be shown by applying bracket multiplication by to (22) $k = n-1, n-2, \dots, 1$ times. Then we find that

$$\sum_{\nu=0}^{n-1} \beta_{\nu}\left\{C, \Omega^{\nu+k}\right\} = \sum_{\nu=0}^{n-k-1} \beta_{\nu}\left\{C, \Omega^{\nu+k}\right\} = 0, \tag{23}$$

and this gives recursively

$$\beta_{o}\left\{C, \Omega^{n-1}\right\} = \beta_{o}\left\{C, \Omega^{n-2}\right\} = \dots = \beta_{o}\,[C, \Omega] = 0. \tag{24}$$

Since $\beta_{o} \neq 0$, combining (22) and (24) we find that $C = 0$, and this finishes the proof of the first part of Theorem IV.

The statement in Theorem IV about (14) being a sufficient condition for (13) to be true is almost trivial. To show that it is not a necessary condition we take $n = 1$ and

$$A(t) = (\cos t - \cos 2t)\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \text{ for } 0 \leqq t \leqq 2\pi$$

$$\tag{25}$$

$$A(t) = \begin{pmatrix} (t-2\pi)^{2} & 0 \\ 0 & (t-2\pi)^{3} \end{pmatrix} \text{ for } t \geqq 2\pi .$$

Clearly,

$$\left[A(t), \int_{o}^{t} A(s)ds\right] \equiv 0 \tag{26}$$

but if $0 < s_{1} < \frac{\pi}{4}$ and if $s_{2} > 2\pi$, then

$$\left[A(s_{1}), A(s_{2})\right] \neq 0. \tag{27}$$

It remains an open question whether (14) is a necessary condition for the vanishing of all Lie-integral functionals of weight $m > n$ if $A(t)$ can be expanded in a power series in t.

VI. Conditions for the Existence of a Solution $Y = \exp\Omega$ for $\dot{Y} = AY$.

We consider again a system of linear differential equations of the first order:

$$\frac{dY}{dt} = A(t)\,Y(t)\,, \tag{1}$$

where Y and A are n by n matrices the elements of which are functions of a parameter t. We assume again that Y(o) is the identity I and that A(t) is continuous in t, although it is well known that the latter condition could be weakened.

If we wish to represent the solution of (1) in the form $Y = \exp\Omega$, the technique based on (III.17) and (III.18) will work for sufficiently small values of $|t|$. Also, it is well known that any preassigned constant matrix Y can be written in the form $\exp\Omega$, if the determinant $|Y|$ of Y is different from zero. From Section V we know that Y is finite and $|Y| \neq 0$ everywhere. Nevertheless, if the function $\Omega(t)$ is assumed to be differentiable, it may not exist everywhere. This can be shown by the following considerations. We may assume that for a fixed value of $t = t_o$ we may start from $t = 0$ and arrive at the value for the solution of (1):

$$Y_o = Y(t_o) = \exp\Omega_o\,, \quad \Omega_o = \Omega(t_o). \tag{2}$$

Let us consider Y_o and Ω_o as points S_y and S_ω respectively of $2n^2$-dimensional space, where the Cartesian coordinates η_ν and ω_ν of these spaces consist of the real and imaginary parts of the n^2 elements of an arbitrary matrix Y or Ω. The formula

$$Y = \exp\Omega \tag{3}$$

defines a mapping of S_ω into S_y such that the coordinates in S_y become entire analytic functions of the coordinates of S_ω. The functional determinant

$$\Delta = \left|\frac{\partial\eta_\nu}{\partial\omega_\mu}\right| \qquad (\nu,\mu = 1,\ldots,2n^2) \tag{4}$$

of this mapping is an analytic function of the ω_μ. If Δ does not vanish at a certain point Ω_o of S_ω, then a neighborhood of Ω_o is mapped continuously with a one-to-one correspondence upon a full neighborhood of Y_o. In this case, the solution $Y = \exp\Omega$ of (1) can be continued beyond the value $t = t_o$ to a value $t_1 > t_o$. If, however, $\Delta = 0$ at $Y = Y_o$, then $dY/dt = AY$ may point towards a part of S_y which is not covered by the map of S_ω in S_y. In this case, $d\Omega/dt$ cannot exist at $\Omega = \Omega_o$. Actually, the question reduces to the problem of solving

$$\left\{\frac{d\Omega}{dt}, \frac{e^{\Omega}-1}{\Omega}\right\} = A \tag{5}$$

with respect to $d\Omega/dt$. If this is possible for any A in the neighborhood (in S_ω) of a point $\Omega = \Omega_o$, and if the result is of the type

$$\frac{d\Omega}{dt} = F(A, \Omega), \tag{6}$$

where the right-hand side in (6) is a matrix depending analytically on the elements of Ω, then (1) has a solution of type (3) in the neighborhood of $Y_o = \exp\Omega_o = \exp\Omega(t_o)$. We shall prove the following result:

Theorem V. The functional determinant Δ defined by (4) does not vanish, and (5) can be solved by an expression (6) in the neighborhood of any point Ω_o in S_ω for an arbitrary A if and only if none of the differences between any two of the eigenvalues of Ω_o equals $2m\pi i$, *where* $m = \pm 1, \pm 2, \ldots$ $m \neq 0$.

Proof: If (3) holds, the determinant $|Y|$ of Y is different from zero. Putting

$$dY \cdot Y^{-1} = dZ, \tag{7}$$

we may compute the determinant which connects the elements of dZ and $d\Omega$. It will differ from Δ only by a power of $|Y|^{-1}$ since the elements of each row of dZ are obtained from the corresponding row of dY by a linear substitution, the matrix of which is the transpose of Y^{-1}. From (III.3) we have

$$dZ = d\Omega - \frac{1}{2!}[d\Omega, \Omega] + \frac{1}{3!}[[d\Omega, \Omega]\Omega] + \ldots . \tag{8}$$

Let

$$dZ = (dz_{\nu,\mu}), \quad d\Omega = (d\omega_{\nu,\mu}), \quad (\nu,\mu = 1,\ldots,n) \tag{9}$$

and assume that

$$\Omega = \Lambda = |\lambda_1, \ldots, \lambda_n| \tag{10}$$

is a diagonal matrix with the numbers $\lambda_1, \ldots, \lambda_n$ in the main diagonal. In this case we have from (8), (9) and (10):

$$(11) \qquad dz_{\nu,\mu} = \frac{\exp(\lambda_\nu - \lambda_\mu) - 1}{\lambda_\nu - \lambda_\mu} \, d\omega_{\nu,\mu} .$$

The n^2 quantities $dz_{\nu,\mu}$ are linear functions of the n^2 quantities $d\omega_{\nu,\mu}$, and the determinant of (11) is

$$(12) \qquad \Delta^* = \prod_{\nu \neq \mu} \frac{\exp(\lambda_\nu - \lambda_\mu) - 1}{\lambda_\nu - \lambda_\mu} ,$$

where the product is extended over $\nu, \mu = 1, \ldots, n$, with $\nu \neq \mu$. Let

$$(13) \qquad (x-\lambda_1)(x-\lambda_2) \ldots (x-\lambda_n) = x^n - s_1 x^{n-1} + s_2 x^{n-2} + \ldots + (-1)^n s_n ,$$

where the s_ν ($\nu = 1, \ldots, n$) are the elementary symmetric functions of the λ_ν. Then Δ^* becomes a function of the s_ν which is analytic and entire in each s_ν. Next, let

$$(14) \qquad \Omega = C \Lambda C^{-1} ,$$

where C is a matrix the determinant of which equals unity. If we introduce

$$(15) \qquad dZ^* = C^{-1} \cdot dZ \cdot C, \quad d\Omega^* = C^{-1} \cdot d\Omega \cdot C ,$$

equation (8) becomes

$$(16) \qquad dZ^* = d\Omega^* - \frac{1}{2!} [d\Omega^*, \Lambda] + \frac{1}{3!} [[d\Omega^*, \Lambda], \Lambda] \mp \ldots$$

and instead of (11) we have

$$(17) \qquad dZ_{\nu,\mu} = \frac{\exp(\lambda_\nu - \lambda_\mu) - 1}{\lambda_\nu - \lambda_\mu} \, d\omega_{\nu,\mu} .$$

Now

$$(dZ^*_{\nu,\mu}) = C^*(dZ_{\nu,\mu}), \quad (d\omega_{\nu,\mu}) = C^*(d\omega_{\nu,\mu}) ,$$

where $C^* = C^{-1} \otimes C$ is the Kronecker product of C^{-1} and C, and where $(dZ_{\nu,\mu})$, etc., stands for the vector with n^2 components $dZ_{\nu,\mu}$. Therefore the $dz_{\nu\mu}$ are again linear functions of the $d\omega_{\nu,\mu}$, where the determinant of the relation is Δ^*. This shows that

$$(18) \qquad \Delta^* = \left| \frac{\partial Z_{\nu,\mu}}{\partial \omega_{\nu,\mu}} \right|$$

whenever Ω can be transformed into diagonal form. Now Δ^* in (12) is a function of the s_ν, which are the coefficients of the characteristic equation for Ω. Since Δ^* must be a continuous function of the $\omega_{\nu,\mu}$, let Δ^* be written as a function of the s_ν. This gives an expression for Δ^* which is valid if Ω has different eigenvalues. But in S_ω the neighborhood of every point Ω_0 contains points corresponding to matrices which have different eigenvalues. Therefore the Δ^* in (18) is given by (12) for all Ω. This proves Theorem V, since Δ^* will vanish if and only if the conditions of Theorem V are satisfied.

VII. The Example of the Ordinary Differential Equation of Second Order

The equation

$$y'' + Q(t)y = 0 \tag{1}$$

may be written as

$$\begin{aligned} y_1' &= y_2 \\ y_2' &= -Qy_1 \end{aligned} \tag{2}$$

and leads to the matrix equation

$$\frac{dY}{dt} = AY , \qquad Y = \begin{pmatrix} y_1 & \eta_1 \\ y_2 & \eta_2 \end{pmatrix}, \quad A = \begin{pmatrix} 0 & 1 \\ -Q & 0 \end{pmatrix}, \tag{3}$$

where y_1, η_1 denote two linearly independent solutions of (1). If we choose these in such a way that

$$\begin{aligned} & y_1(0) = 1 \quad y_2(0) = y_1'(0) = 0 \\ & \eta_1(0) = 0, \ \eta_2(0) = \eta_1'(0) = 1 , \end{aligned} \tag{4}$$

then $Y(0) = I$ and the determinant $|Y|$ of Y equals unity for all values of t. Therefore, if $Y = \exp \Omega$, the trace of Ω vanishes and we may put Ω into the form

$$\Omega = \begin{pmatrix} \omega & \phi \\ \Upsilon & -\omega \end{pmatrix} . \tag{5}$$

If we introduce

$$\Delta = \sqrt{\omega^2 + \phi \Upsilon} , \tag{6}$$

we find after some calculation:

$$Y = \frac{(\sinh \Delta)}{\Delta} \Omega + (\cosh \Delta) I . \tag{7}$$

Let

$$\theta = y_1 + \eta_1 . \tag{8}$$

be the trace of Y. Then we have from (7)

$$(9) \qquad \theta = 2 \cosh \Delta$$

(since the trace of Ω vanishes). Actually, 2Δ is the difference between the eigenvalues λ_1, λ_2 of Ω. If this difference is a multiple of $2\pi i$ but different from zero, then $\lambda_1 \neq \lambda_2$ since $\lambda_1 + \lambda_2 = 0$. In this case, Ω can be transformed into the diagonal form. But then, although both the eigenvalues of Y equal +1 or -1, it is possible that Y is not a multiple of the unit matrix. Indeed, if we combine (7) and (9) we find

$$(10) \qquad \Omega = (2Y - \theta I) \frac{\Delta}{\sqrt{\theta^2 - 4}}$$

which shows that for $\Delta = 2n\pi i$ we have $\theta^2 = 4$; therefore if $\Delta \neq 0$ and is finite, $2Y - \theta I = 0$. In order to discuss at least one case completely we prove

Theorem VI. *Let* $Q(t) > 0$ *for all* $t > 0$. *Then a solution* Y of (3), *with the initial condition* $Y(0) = I$, *has a representation*

$$Y(t) = \exp \Omega(t), \qquad \Omega(0) = 0, \quad (t \geqq 0)$$

with a differentiable $\Omega(t)$, *if and only if*

$$(11) \qquad (\text{trace } Y)^2 \leqq 4$$

for $t \geqq 0$.

Proof: If $Y = \pm I$, for $t = t_o$ we have $y_1 = \pm 1$, $\eta_1' = \pm 1$, $y_1' = 0$, $\eta_1 = 0$, and

$$(12) \qquad \frac{d\theta}{dt} = y_1' + \eta_1'' = y_1' - Q\eta_1 = 0$$

$$(13) \qquad \frac{d^2\theta}{dt^2} = y_1'' - Q\eta_2 - Q'\eta_1 = -Q\theta - Q'\eta_1 = \mp 2Q .$$

Therefore if $Q > 0$ then $\theta^2 < 4$ for $t = t_o + \varepsilon$, if ε is positive and sufficiently small. By differentiating (9), we find $\Delta = iD$, where $\cos D = \pm 1$, $D' = 0$, and $D'' = \pm \frac{1}{2} \theta''$ is positive. Therefore, D increases. Now we let t increase. We arrive at a point where $\theta' = 0$, $\theta^2 < 4$ or at a point t_1 where $\theta^2 = 4$; D has been increasing in the interval $t_0 \leqq t \leqq t_1$. The possibility that θ' vanishes is excluded because $y_1\eta_1' - y_1'\eta_1 = 1$, and if $\theta^2 < 4$, $\theta' = 0$, then

(14) $$|y_1\eta_1'| = |y_1\eta_2| < 1, \quad y_1\eta_1' - y_1'\eta_1 = y_1\eta_1' - Q\eta_1^2 < 1,$$

which involves a contradiction. Therefore, D must increase until θ^2 reaches the value 4 again. But since we started from $D = 0$, we have arrived at a value of $D = n\pi$, $n > 0$, and therefore, if Ω shall exist, Y must be the identity. This proves that the condition $\theta^2 \leqq 4$ is a necessary one. But it is also sufficient. We prove this as follows; If $\theta^2 \leqq 4$ for $t \geqq 0$ we have $d\theta/dt = 0$ if $\theta = \pm 2$, and therefore:

(15) $$y_1 + \eta_1' = \pm 2, \quad y_1' - Q\eta_1 = 0, \quad y_1\eta_1' - y_1'\eta_1 = 1.$$

But then we have $|y_1\eta_1'| \leqq 1$, and from (14) we see that $y_1\eta_1' = 1$, $Q\eta_1^2 = 0$, which implies $\eta_1 = 0$. Then $y_1' = 0$ and from this we have $y_1 = \eta_1' = \pm 1$. Therefore $Y = I$ whenever $\cosh\Delta = \pm 1$, $\Delta \neq 0$. Then Ω is completely determined for $\sinh\Delta \neq 0$ by (10) and (9) and for $\sinh\Delta = 0$, $\Delta \neq 0$ by the result obtained from differentiating (10), namely

(16) $$\Delta Y' = (\Delta' \cosh\Delta)\Omega .$$

By substituting the explicit expression (5) for Ω and the product AY for Y' we find from (16) and (6)

(17) $$\phi\Delta' = \Delta, \quad \psi\Delta' = -Q\Delta, \quad \omega = 0, \quad \phi\psi = \Delta^2,$$

where $\Delta = n\pi i$, $n = \pm 1, \pm 2, \ldots$ From (17) we can derive

(18) $$\Delta' = \pm\sqrt{-Q} ,$$

where the sign of the square root in (18) is determined by the postulate of the differentiability of Δ. Now Ω is completely and uniquely determined everywhere, and this completes the proof of Theorem VI.

References

[1] K. O. Friedrichs. *Mathematical aspects of the quantum theory of fields, Part V.*, Communications on Pure and Applied Mathematics, Vol.6, pp. 1-72, (1953).

[2] Herbert B. Keller and Joseph B. Keller, *On systems of linear ordinary differential equations*, Research Report EM-33, New York University, Washington Square College of Arts and Science, Mathematics Research group.

[3] H. F. Baker, *On the integration of linear differential equations*, Proceedings of the London Mathematical Society, Vol. 35, pp.333-374, (1903); Vol. 34, 347-360, (1902); Second Series, Vol.2, 293-296, (1904).

[4] H. F. Baker, *Alternants and continuous groups*, Proceedings of the London Mathematical Society Second Series, Vol. 3, pp.24-47, (1904).

[5] F. Hausdorff, *Die symbolische Exponentialformel in der Gruppentheorie*, Berichte der Saechsischen Akademie der Wissenschaften (Math.Phys. Klasse), Leipzig, Vol. 58, 19-48, (1906).

[6] R. P. Feynman, *An operator calculus having applications in quantum electrodynamics*, Physical Review Vol. 84, No. 1, pp.108-128, (1951).

[7] W. Magnus, *Ueber Beziehungen zwischen hoeheren Kommutatoren*, Journal f. d. reine u. angewandte Mathematik, Vol. 177, pp. 105-115, (1937).

[8] E. Witt, *Treue Darstellung Liescher Ringe*, Journal f.d. reine u. angewandte Mathematik, Vol. 177, pp. 152-160, (1937).

[9] W. Magnus, *A connection between the Baker-Hausdorff formula and a problem of Burnside*, Annals of Mathematics, Vol. 52, pp. 111-126, (1950).

[10] G. Falk, *Konstanzelemente in Ringens mit Differentiation*, Mathematische Annalen, Vol. 124, pp. 182-186, (1952).

[11] H. Zassenhaus, Unpublished.

[12] E. B. Dynkin, *Calculation of the coefficients in the Campbell-Hausdorff formula*, Doklady Akad. Nauk, USSR (N.S.), Vol. 57, 323-326, (1947) (Russian).

[13] W. Specht, *Die linearen Beziehungen zwischen hoeheren Kommutatoren*, Mathematische Zeitschrift, Vol. 51, pp. 367-376, (1948).

[14] F. Wever, *Operatoren in Lieschen Ringen*, Journal f. d. reine u. angewandte Mathematik, Vol. 187, pp. 44-55, (1947).

Date Due

~~JAN 2 '57~~
~~FEB 27 '58~~
~~MAR 5 '59~~
~~JUN 7 [illegible]~~
~~5/9/62~~
~~NOV 26 '82~~
~~DEC 9 '84~~
~~out 6-4-75~~
~~ILL Sears due 6-25-75~~
MAY 13 1992

PRINTED IN U.S.A.

NYU, ~~Math.res.group~~ Res.rep.
AUTHOR BR-3
~~Wilhelm~~ Magnus
TITLE
Algebraic aspects...
DATE

BR-3

NEW YORK UNIVERSITY
INSTITUTE OF MATHEMATICAL SCIENCES
LIBRARY
25 Waverly Place, New York 3, N. Y.

MAGNUS

ALGEBRAIC ASPECTS...

NEW YORK UNIVERSITY
INSTITUTE OF MATHEMATICAL SCIENCES
LIBRARY
4 Washington Place, New York 3, N. Y.